We claim:

1. A method for automatically evaluating an essay to detect at least one writing style error, comprising:

electronically receiving an essay on a computer system;

assigning a feature value for each of one or more features for one or more text segments in the essay, wherein the feature values are automatically calculated by the computer system;

storing the feature values for the one or more text segments on a data storage device accessible by the computer system;

comparing the feature values for each one or more text segments with a model configured to identify at least one writing style error, wherein the model is based on at least one human evaluated essay; and

using the results of the comparison to the model to identify writing style errors in the essay.

2. The method according to claim 1 wherein the writing style error is the overly repetitive use of one or more text segments.

3. The method of claim 1 wherein the text segment comprises a word.

4. The method of claim 1 wherein the comparison step comprises extracting patterns from the feature values, wherein the patterns are based on the presence of absence of features associated with each word in the essay.

5. The method of claim 1, wherein the function words of the essay are not considered by the computer system in determining the feature values.

6. The method of claim 1 wherein the feature values comprise the total number of times the evaluated text segment occurs in the essay.

7. The method of claim 1 wherein the feature values comprise the ratio of the evaluated text segment occurrences in the essay to the total number of text segments in the essay.

8. The method of claim 1 wherein the feature values comprise the average, over all paragraphs of the essay, of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay, over the total number of text segments in the paragraph.

9. The method of claim 1 wherein the feature values comprise the largest value of the ratio, of the number times the evaluated text segment occurs in a paragraph of the essay, over the total number of text segments in the paragraph, wherein the ratio is calculated for each paragraph in the essay.

10. The method of claim 1 wherein the feature values comprise the length, measured in characters, of the text segment.

11. The method of claim 1 wherein the feature values comprise a value indicating whether the text segment includes a pronoun.

12. The method of claim 1 wherein the feature values comprise a value representing the interval distance between consecutive text segment occurrences.

13. The method of claim 12 wherein the distance is determined by calculating the number of intervening words.

14. The method of claim 12 wherein the distance is determined by calculating the number of intervening characters.

15. The method of claim 1 wherein the model is generated using a machine learning tool.

16. A system for automatically evaluating an essay to detect at least one writing style error, comprising:

a computer system configured to electronically receive an essay;

a feature extractor configured to assign a feature value for each of one or more features for one or more text segments in the essay;

a data storage device, connected to the computer system, configured to store the feature values for the one or more text segments;

a feature analyzer configured to evaluate the essay for at least one writing style error by comparing the feature values for each one or more text segments with a model; and

a display for presenting the evaluated essay.

17. The system of claim 16 wherein the writing style error is the overly repetitive use of one or more text segments.

18. The system of claim 16 wherein the text segment comprises a word.

19. The system of claim 16, further comprising:

an annotator configured to annotate the essay to identify the one or more writing style errors.

20. The system of claim 16 wherein the feature extractor comprises an occurrences calculator configured to generate a value representing the total number of times the text segment occurs in the essay.

21. The system of claim 16 wherein the feature extractor comprises an essay ratio calculator configured to generate a value representing the ratio of the number of times the evaluated text segment occurs in the essay to the total number of text segments in the essay.

22. The system of claim 16 wherein the feature extractor comprises an average paragraph ratio calculator configured to generate a value representing the average over all paragraphs in the essay of the ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

23. The system of claim 16 wherein the feature extractor comprises a highest paragraph ratio calculator configured to generate a value representing the largest ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

24. The system of claim 16 wherein the feature extractor comprises a length calculator configured to generate a value representing the length, measured in characters, of the text segment.

25. The system of claim 16 wherein the feature extractor comprises an identifier to determine whether the text segment includes a pronoun.

26. The system of claim 16 wherein the feature extractor comprises a distance calculator configure to generate a value representing the distance between consecutive text segment occurrences.

27. The system of claim 26 wherein the distance between consecutive text segment occurrences is measured in words.

28. The system of claim 26 wherein the distance between consecutive text segment occurrences is measured in characters.

29. The system of claim 16 comprising a machine learning tool to generate the model.

30. The system of claim 16 where in the model is generated using at least one human evaluated essay.

31. A method for generating a model for determining overly repetitive text segment use, comprising:

electronically receiving training data on a computer system wherein the training data comprises an essay annotated to identify one or more text segments used in an overly repetitive manner;

assigning a feature value for each of one or more features for each text segment in the essay, wherein the feature values are automatically calculated by the computer system;

assigning an indicator value for each text segment in the essay, wherein the indicator value is set at a first value and if the text segment has been used in an overly repetitive manner;

storing the feature values and the indicator value for each text segment in the essay in a data storage device accessible by the computer system; and

creating a model for overly repetitive use of the one or more text segments in the essay by identifying patterns in the feature values wherein the patterns are identified by a machine learning tool.

32. The method of claim 31 wherein the text segment comprises a word.

33. The method of claim 31 wherein the annotations are manual markings.

34. The method of claim 31, wherein the function words of the essay are not considered by the computer system in calculating the feature values.

35. The method of claim 31 wherein the feature values comprise the total number of times the evaluated text segment occurs in the essay.

36. The method of claim 31 wherein the feature values comprise the ratio of the evaluated text segment occurrences in the essay to the total number of text segments in the essay.

37. The method of claim 31 wherein the feature values comprise the average over all paragraphs of the essay of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

38. The method of claim 31 wherein the feature values comprisethe largest value of the ratio of the number times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph, wherein the ratio is calculated for each paragraph in the essay.

39. The method of claim 31 wherein the feature values comprise the length, measured in characters, of the text segment.

40. The method of claim 31 wherein the feature values comprise a value indicating whether the text segment includes a pronoun.

41. The method of claim 31 wherein the feature values comprise a value representing the interval distance between consecutive text segment occurrences.

42. The method of claim 41 wherein the distance is determined by calculating the number of intervening words.

43. The method of claim 41 wherein the distance is determined by calculating the number of intervening characters.

44. A system for generating a model useful in determining overly repetitive text segment use, comprising:

a computer system configured to receive training data, wherein the training data comprises an essay annotated to identify one or more text segments used in an overly repetitive manner;

a feature extractor configured to calculate a feature value for each of one or more features for each text segment in the essay and to assign an indicator value for each text segment in the annotated essay, wherein the indicator value indicates whether the text segment has been used in an overly repetitive manner;

a data storage device configured to store the feature values and the indicator value for each text segment in the essay;

a machine learning tool configured to analyze the features to identify patterns; and

a model builder to create a model for overly repetitive use of the text segments, wherein the model is constructed from the identified patterns.

45. The system of claim 44 wherein the annotated essays are manually marked.

46. The system of claim 44 wherein the feature extractor comprises an occurrences calculator configured to generate a value representing the total number of times the text segment occurs in the essay.

47. The system of claim 44 wherein the feature extractor comprises an essay ratio calculator configured to generate a value representing the ratio of the number of times the evaluated text segment occurs in the essay to the total number of text segments in the essay.

48. The system of claim 44 wherein the feature extractor comprises an average paragraph ratio calculator configured to generate a value representing the average over all paragraphs in the essay of the ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

49. The system of claim 44 wherein the feature extractor comprises a highest paragraph ratio calculator configured to generate a value representing the largest ratio of the number of times the evaluated text segment occurs in a paragraph of the essay over the total number of text segments in the paragraph.

50. The system of claim 44 wherein the feature extractor comprises a length calculator configured to generate a value representing the length, measured in characters, of the text segment.

51. The system of claim 44 wherein the feature extractor comprises an identifier to determine whether the text segment includes a pronoun.

52. The system of claim 44 wherein the feature extractor comprises a distance calculator configure to generate a value representing the distance between consecutive text segment occurrences.

53. The system of claim 52 wherein the distance between consecutive text segment occurrences is measured in words.

54. The system of claim 52 wherein the distance between consecutive text segment occurrences is measured in characters.